

# 浅谈主成分分析与因子分析方法的联系与区别

滕 达

(吉林工商学院会计分院,吉林 长春 130062)

摘 要: 介于主成分分析方法和因子分析方法均是通过“降维”的思想解决多变量、大样本的问题,两者在实际使用过程中既有联系,又有区别。故本文旨在通过分析主成分分析和因子分析方法的内涵,对两者异同进行比较分析,以便更进一步了解两种方法的各自特点,更好的用其解决科学研究中的相关问题。

关键词: 主成分分析 因子分析 联系 区别

中图分类号: C81

文献标识码: A

## 1 问题的提出

在现实生活或科学研究过程中,影响某一事物的特征或该事物发展规律的因素是多元化的,我们在对这些影响因素对于事物的影响进行研究过程中,该事物的某一特征作为统计学意义上的因变量,而影响因素则作为自变量。为了更加全面的对事物的特征或发展规律进行反映,需要综合与其相关各种影响因素进行评价,即在研究过程中对于影响事物特征或发展规律的因素需要更多的引入,对其进行综合分析和评价。然而,多变量大样本资料尽管可以对事物特征或发展规律提供更加全面的信息,但同时带来了多重共线性等问题,使得影响因素所反映的信息重复,影响统计结果的真实性和科学性。对此,降维思想成为解决这一问题的有效方式。主成分分析和因子分析方法都是运用降维的思想,将多变量信息归纳为少数几个相互无关的综合变量以反映原来数据的大部分信息。

近年来,主成分分析和因子分析方法作为一种统计分析方法在科学研究中的应用十分广泛,运用其进行多变量分析的学术文献越来越多。然而,在实际使用过程中,常常出现一些将两种方法进行混淆的错误,由此产生的统计分析结果在科学性上大打折扣。因子分析方法是主成分分析方法的推广和发展,两种方法之间既存在共同之处,也有着显著的差别,有必要对两种方法之间的联系和区别进行严格区分,并针对实际问题选择恰当的分析方法。

## 2 两者的联系与区别

### 2.1 两者的联系

主成分分析和因子分析方法都属于多元统计分析中处理降维的统计方法。在数理统计的基本原理上,两者都是基于多变量的相关系数矩阵,在确保较少信息缺失的前提下(一般小于或等于 15%),用少数几个不相关综合变量概括多个变量的信息(多个变量之间存在较强的相关性)。即用少数不相关的综合变量尽可能全面的反映多个原始变量的信息,消除了原始变量的相关性,可信度得到提高,统计结果可以有效地解释现实问题。需要注意的是,两种方法产生的新的变量(因子)不是原始变量筛选后的剩余变量,而是综合所有变量信息后的新变量。其中,在主成分分析过程中,新变量是原始变量的线性组合,即将多个原始变量经过线性(坐标)变换得到新的变量。在因子分析过程中,新变量则是通过原始变量之间的复杂关系对原始变量进行分解,得到公共因子和特殊因子。其中公共因子是所有原始变量中所共同具有的特征,而特殊因子则是原始变量所特有的部分。两种方法下得到的主成分变量与因子变量在数量上显著少于原始变量,起到了降维的作用,也提高了数据有效利用程度。

### 2.2 两者的区别

#### 2.2.1 基本概念不同

主成分分析法是将多个指标转化为少数彼此不相关的综合指标(即主成分)的统计方法。而因子分析法是主成分分析法的推广和发展,它也是将具有错综复杂关系的变量综合为数量较少的几个因子,再根据不同因子还可以对变量进行分类。同时重塑原始变量与因子之间的相互关系。

#### 2.2.2 基本原理不同

主成分分析方法旨在通过方差-协方差矩阵将多个原始变量通过多次线性变换得到少数几个主成分(新的变量),这些主成分变量能够反映原始变量尽可能多的信息(一般大于或者等于 85%为通过标准),并且它们之间不相关。从数理上讲,主成分分析法是一种矩阵变换的方法,即将给定的变量(原始变量)通过多次线性变换,转换成一组彼此不相关的变量,在这个过程中,变量的方差之和保持不变,方差最大的作为第一主成分变量,以此类推,得到数量较少的、可以涵盖大部分原始变量信息几个主成分。从这个意义上讲,主成分分析法是作为因子分析的一种方式。

而因子分析法则通过原始变量的相关系数矩阵将变量进行分组,分组的原则是将相关性较高的变量置于一组中,但组与组之间的变量相关性较低。这样各组变量代表一个基本要素(公共因子),所研究的问题可以分解为少数几个公共因子的线性函数与特殊因子之和。可见,因子分析法下的新变量是对原始变量进行分解得到,而不是原始变量的线性组合。具体而言,就是通过获取原始变量中可测量的、具有一定相关性的统计指标测定各个因子的状态。从该意义上讲,因子分析只能解释变量的部分变异,而主成分分析法则解释了所有变异。

#### 2.2.3 数据处理过程不同

在消除量纲和数量级的处理上,主成分分析通常需要对原始数据进行标准化处理,将原始数据转换为均值为 0、方差为 1 的标准化数据。而因子分析法对此则要求不高,这是因为因子分析法本身可以通过加权最小二乘法、主成分法等求解因子变量,在这个过程中,因子是原始变量内部分解的结果,与原始变量是否同量纲关系不大。只有通过主成分法确认因子变量时,需要对原始数据进行无量纲化处理。另外一点,主成分分析法下的新变量(主成分)是通过原始变量的多次线性组合后得到的,这个过程本身具有可逆性;而在因子分析法下,因子分析中的载荷矩阵是不可逆的,只能通过可观测的原变量去估计不可观测的公共因子。

此外,主成分分析法主要侧重于变量的信息贡献能力,而因子分析法则侧重于因子的可解释性。

2.2.4 统计软件实现过程不同 - 以 SPSS 为例

在利用统计软件 SPSS 进行主成分分析时,其基本步骤大致为:(1)对原始数据进行标准化处理;(2)选择“分析(Analyze)-数据提取(Data Reduction)-成分分析(Factor Analyze)”打开主成分分析对话框,在“数据描述(Descriptives):相关系数矩阵(Correlation Matrix)”框中选系数(Coefficients);“统计(Statistics)”框中选初始解(Initial solution) Step4:“提取(Extraction):方式(Method)”框中选主成分(Principal components);“分析(Analyze)”框中选相关系数矩阵(Correlation matrix);“显示(Display)”框中选未经旋转的因子载荷(Unrotated factor solution);“提取(Extract)”框中选特征值(Eigenvalues) Step5:结果显示在 Output 中:提取方差总合计(Total Variance Explained)中主成分的累计贡献率大于等于 85%的主成分个数;“Component Matrix”中第 i 个主成分的列向量除以相应特征根的平方根后就得到这个主成分的变量系数向量,可以利用“Transform-compute”来实现 Step6:写出主成分表达式及主成分命名。

在利用统计软件 SPSS 进行因子分析时,其基本步骤为:(1)-(4)同主成分分析的实现过程;(5)“旋转(Rotation):方式(Method)”框中选最大方差法(Varimax);“显示(Display)”框中选择旋转的因子载荷阵(Rotated solution);(6)“得分(Scores)”框中变量形式保存(Save as variables);“方法(Method)”框中选回归(Regression);(7)结果同样显示在 Output 中,提取方差总合计(Total Variance Explained)中主成分的累计贡献率大于等于 85%的因子个数,并对各个因子进行命名,并根据因子得分函数对因变量进行排序。

### 结束语

目前,主成分分析法和因子分析法在科学研究中应用十分广泛,在处理多变量、大样本的统计问题上优势十分明显。因此,广受学术界学者们的青睐。但是,在现实中,因为两种方法之间存在着诸多内在的联系,不少学者在运用这两种方法解决问题过程中常常出现混淆使用的现象,统计软件输出的结果也很难解释,科学性和可信性大打折扣。因此,明确两种方法之间的联系和区别,正确、恰当的使用这两种方法,对于学术研究十分关键。本文正基于此,探讨了两种方法之间的联系和区别,以期抛砖引玉。

### 参考文献

- [1] 魏艳华,王丙参,田玉柱.主成分分析与因子分析的比较研究,天水师范学院学报,2009 年第 3 期.
- [2] 景慧丽.主成分分析和因子分析比较.商业文化,2008 年第 8 期.
- [3] 马娟,杨益民.主成分分析与因子分析之比较及实证分析.市场研究,2007 年第 3 期.